DNA Mark-Recapture
Lessons Learned and New Developments

John Boulanger

Integrated Ecological Research
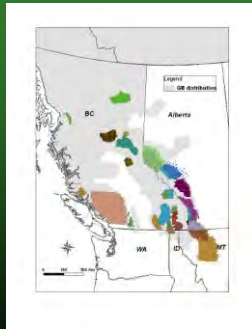

Outline

- A historical review…
- Estimating population size and density
  - HS only
  - Multiple data sources
- Estimating trend

DNA hair-snag mark-recapture projects

- Over 20 projects conducted since 1996

- All grid-based projects have been 1 year projects to estimate population size

- 2 months duration

Back in 1996…..

- Genotyping from hair a novel approach
- Hair snag sites the main way to collect DNA
- Program CAPTURE…

In 2012….

- Rub trees, scats to collect DNA
- Program MARK
  - Use of covariates
  - Multiple models to estimate trend and demography
- Program DENSITY
 - Spatially explicit methods

---

Optimizing sampling design

Study design → Field sampling → Genetic analysis → Mark-recapture estimation

Mark-recapture analysis should *complement* previous links

Keep analyses "close" to the data

The "target" population

- All mark-recapture methods assume that every bear has an non-zero detection rate.
- But, bears can have unequal detection rates*
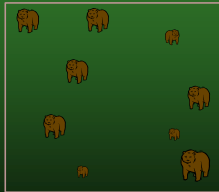
The "target" population

- If some bears are "invisible" to sampling then N estimates will be biased
- Trend estimates would assume "invisible bears" randomly enter/leave the population
- An *index* of population size
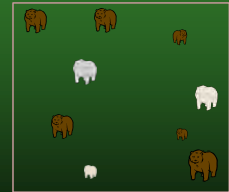
Challenges of DNA-based Population Estimation
Bear densities are sparse and they move a lot!

- Population closure
- Capture probability variation
- Obtaining adequate sample sizes (**N** and **p**)

- *How can we optimize sampling to reduce costs while meeting the above objectives?*

Population Closure

- Causes N to be overestimated
- "Superpopulation" N* (White 1996)
  - N unbiased if "random movement"
  - Sampling area undefined
  - D estimates biased
- Average N-density

Sampling grid

Superpopulation
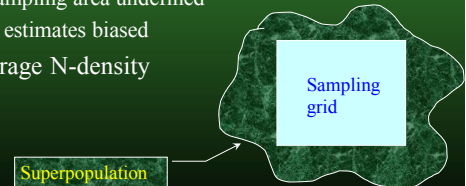
Closure: Spatial and temporal scale of sampling

Log (density)
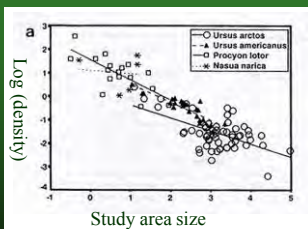
Study area size

Smallwood and Schonewald (1996)

- Difficult to scale study area to population
- Attributes of this problem
  - Study area selection
  - Topography
  - Habitat
- This is an issue with *any method* used to estimate N

Effects of closure violation
Lowered detection probabilities

- Closure violation reduces detection probabilities

- Decreased estimate precision

- Difficult to detect heterogeneity
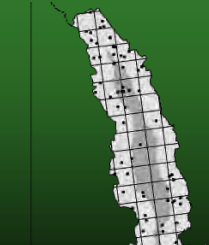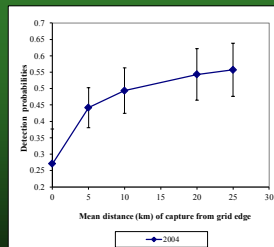
- 65 bears with 56 detected once

Closure--Analysis strategies

→Covariate approach

→Mean detection location for each DNA bear

→Distance from grid edge
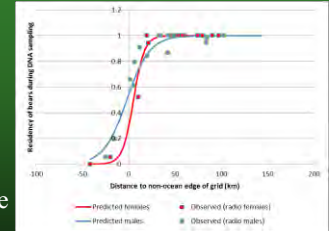
→Covariate for p in Huggins N model

CJZ (2001)

Closure--Analysis strategies to estimate density

→Program MARK

→Radioed bears residency
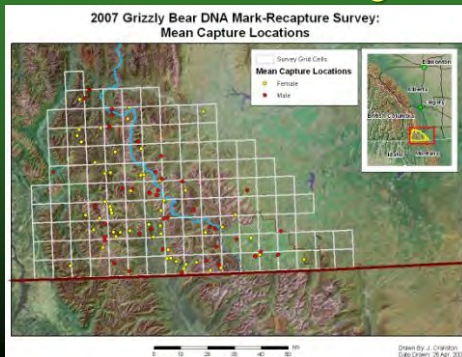
→Mean residency as a function of distance from edge

→Assign to DNA bears

$$\widehat{N} = \sum_{i=1}^{M_{t+1}} \frac{\tilde{r}}{p*}$$

Ivan (2008)

Uneven distribution across grids

2007 Grizzly Bear DNA Mark-Recapture Survey: Mean Capture Locations

Program DENSITY

→SECR methods provide a new way to estimate density etc....

→SECR does make more assumptions about the data

- Circular home range/detection functions
- Poisson distribution of HR centers (covariates)
- Permeability of grid edge

→Simulations provide a way to test robustness to these assumptions

Capture probability (p) variation

→Three types

- *time*-p for population changes each session
- *behavior*-p changes after initial capture (to c)
- *heterogeneity*-each bear has unique p
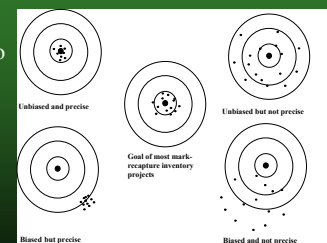 - Sex, previous collaring history

→Genetic

- Shadow effects, allelic drop out etc…...
- RIGOROUS LAB TECHNIQUE ESSENTIAL
- Roon et al (2005) simulations suggest minimal problem for HAIR sampling

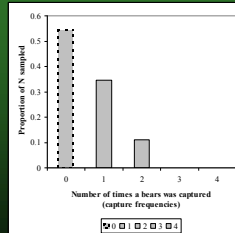$$\widehat{N} = \frac{M}{\hat{p}}$$

Heterogeneity

→Causes

- capture probability to be overestimated
- population size to be underestimated
- variance of population size to be underestimated

Heterogeneity and sparse data

- When capture probabilities are low it is not possible to detect or efficiently model heterogeneity
- Erroneous model selection
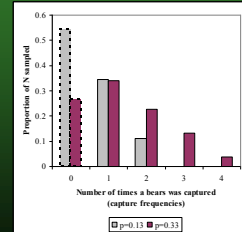- Difficult to conduct >5 sessions

Heterogeneity and sparse data

- When capture probabilities are low it is not possible to detect or efficiently model heterogeneity
- Erroneous model selection
- Difficult to conduct >5 sessions

Heterogeneity and sparse data

- When capture probabilities are low (<0.2) it is not possible to detect or efficiently model heterogeneity
- Erroneous model selection
- Difficult to conduct >5 sessions
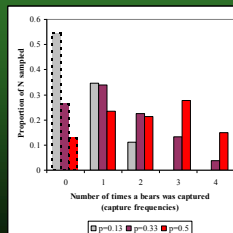
Study design
HS sites are not 100% efficient

- GPS bears movements relative to DNA sites
- 63% of bears that encounter sites are genotyped (1999)
- Previously live captured bears have lower detection probabilities (JWM 2008)
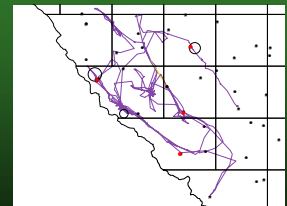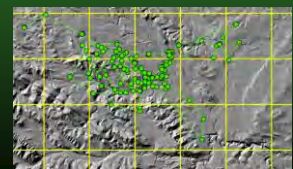

J. Mammalogy (2004)

Field sampling design
Increase capture probabilities

- Capture probability increased from 0.1 to 0.45 due to:
  - Better bait (fish/blood liquid lure)
  - Greater success in genotyping samples ("the Paetkau effect")
  - Site selection
  - Minimize closure
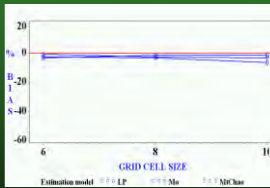
Heterogeneity:
Do we need to move sites?

- Female spring home ranges
  - British Columbia: 50 (25-155) km^2
  - Alberta: 300 (50-1500) km^2
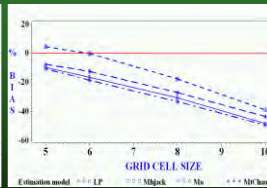
- Female movements
- 7x7 km cell

Simulation Results

Traps moved:

Traps not moved:

GRID CELL SIZE

Estimation model — LP — Mo — MtChao

GRID CELL SIZE

Estimation model — LP — Mbjack — Mo — MtChao

Study design tradeoffs

- Larger grid-move sites
  - Maximizes N
  - Reduces p
  - Reduces closure violation
  - Costs more
- Smaller grid-do not move sites
  - Reduces N
  - Enhances p
  - Increased risk of closure violation
  - Costs less
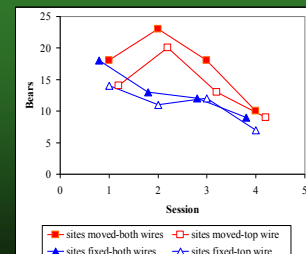

A designed study

- Sampled for 4 sessions with sites moved
- Kept sampling the first site for the remaining 3 sessions
- Created
  - 1 "Moved site" data set
  - 1 "Fixed site" data set
- Double wire vs single wire

Results: Number of bears captured

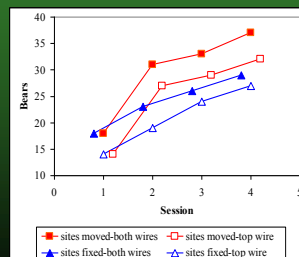- More bears captured with sites moved and double wire sampling
- Same bait site density

Cumulative individuals caught

- More individuals captured with sites moved and double wire
- Similar CPUE fixed-moved site designs
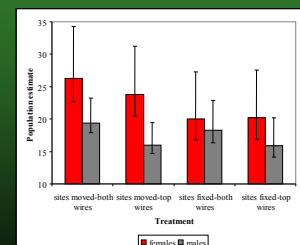

Population estimates

- Sites moved -both wires design had
  - Female estimates 25% higher (vs. fixed)
 - Less difference between male estimates

Ursus (2006)

Conclusions: moving vs fixing sites

- Moving sites
  - captures more bears
  - Better estimates for females
- Fixing sites requires *smaller grid cell size*
  - Unbiased estimates still possible with fixed site designs (5x5 km for grizzly bears)
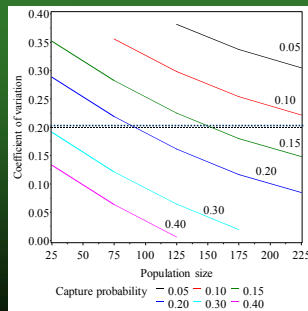- Could SECR minimize bias with fixed sites?

Sample size:
Detection probabilities and population size
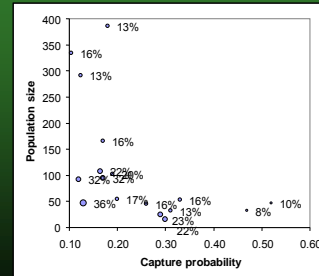
Sample size needed for precise estimates (CV<20%)

- Polynomial regression of CV, N, and p
- Levels of N and p needed to obtain CV
- Stand-alone projects using CAPTURE M_h Chao
  - 4-5 sessions
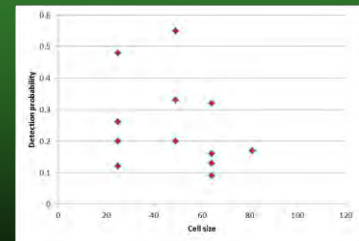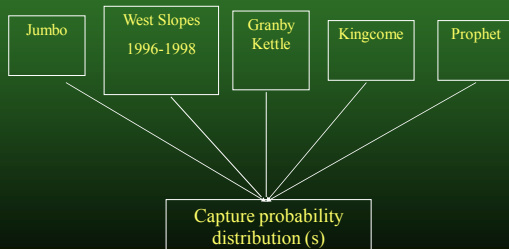

Proctor et al. Ursus (2011)

Sample size
Cells size vs detection probability

- 49 km^2 sites moved and 25 km^2 sites not moved works well

Mitigating sparse data- Meta-analysis-pooling data

Jumbo | West Slopes 1996-1998 | Granby Kettle | Kingcome | Prophet

Capture probability distribution (s)

MARK vs CAPTURE estimates


CAPTURE Mh   MARK Mh mixture
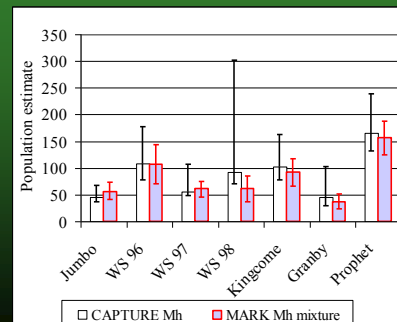
Will a meta analysis approach work for you?

- Standardization of design ESSENTIAL
  - Grid cell size, bait types, trap placement
  - Synchronized timing of surveys
  - Record covariates
  - Estimate closure violation

Ursus (2002)

Multiple data sources
A newer way to mitigate heterogeneity

- Bears use rub trees during hair snag sampling
- Can we combine these two data sources to get better estimates?
- Cheap way to improve estimates
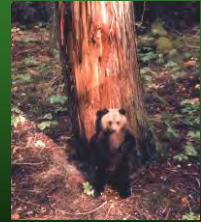

Even cubs use rub trees!

Estimation methods
Hair Snag-Rub Tree (HSRT)

- Lincoln Petersen method-pool data
  - Session 1- individuals id'd with hair snags
  - Session 2-individuals id'd with rub trees
- Program MARK
  - Session 1-5- capture histories hair snags
    - For example-0100
  - Sessions 6-10-capture histories rub trees
    - For example-10000

Comparison of Population Size Estimates
HSRT vs HS vs RT

- Multiple data sources (rub tree and hair snags) produce robust estimates
- Do rub trees sample the entire "target" female population?
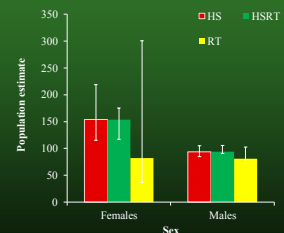
Simulation evaluation

- Key assumption--capture probabilities from hair snags/rub trees cannot be correlated
- Does this matter?
  - Monte Carlo simulation
  - Correlation- r -1 to 1
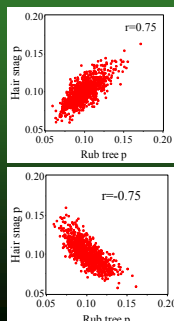
Results
Sparse data-Capture probability=0.1

- Correlation causes varying levels of bias dependent on heterogeneity
- Decreased confidence interval coverage
- Better performance at higher detection rates

7

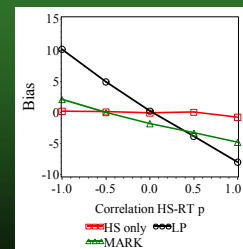## Conclusions: Multiple data sources

- Combining data sources can increase precision assuming
  - Minimal correlation's
  - Capture probabilities >0 from 1 data source
  - Data collected as sessions
- Increasing capture probabilities boosts robustness
- Add in management bears as another session

Kendall et al 2008, 2009 (JWM)
Ecological Applications, (2008)

Estimating population size Summary

- Critical study design strategies
  - Grid placement
  - Optimized lures/site selection
  - Cell size vs moving sites
  - Multiple data sources
- Analysis strategies
  - Meta analyses
- No such thing as a "free lunch"!

Estimating trend:
Advances in mark-recapture estimation

- Evolution from estimation of N to testing of hypotheses about demography and trend
  - Use of covariates
- Flexibility in model fitting
  - Year-specific estimates
  - Pooled estimates

Program MARK
You can obtain context-sensitive help with the F1 key,
and can investigate objects with the Shift-F1 key.
See the Help menu for known problems.

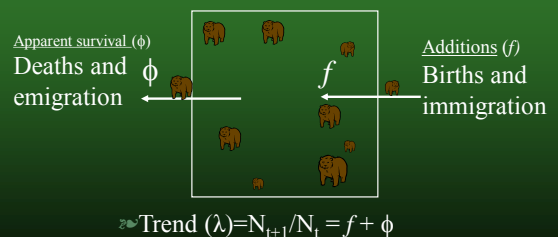## DNA vs Radio collar sampling

- DNA sampling
  - Many bears
  - Less information/bear
- Radio collars
  - Less bears
  - More information/bear
- _Both_ are samples of a population
- Best method *depends on objectives* (Radium workshop)
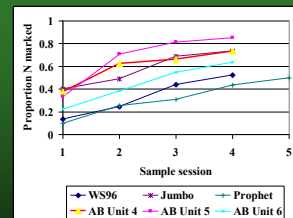
DNA-based monitoring
Multi-year sampling

- Tracking bear fates over time
- The encounter history
  - 100000000--likely bear not on study area
  - 100100101-likely bear on study area
  - 000000001-likely bear immigrated/born
- Utilize full information in data set
  - Different than tracking N estimates over time

Pradel Model
Local Demography

<u>Apparent survival</u> (ϕ)
Deaths and ϕ
emigration

f

<u>Additions</u> (*f*)
Births and
immigration

- Trend (λ)=$N_{t+1}/N_t = f + \phi$
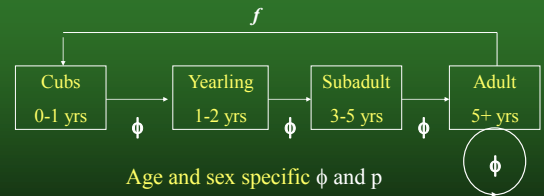
Pradel Model

- Assumes
  - Study area does not change in size
  - Minimal behavioral response
  - Robust to heterogeneity* (Hines and Nichols 2002)
- Demographic assumptions $\lambda = \phi + f$
  - ϕ constant-
  - f constant-
  - all constant-

Monte Carlo simulation

$$f$$

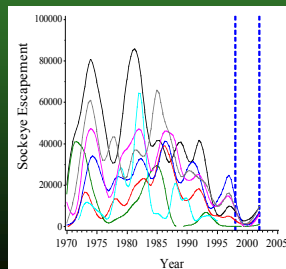| Cubs 0-1 yrs	ϕ	Yearling 1-2 yrs	ϕ	Subadult 3-5 yrs	ϕ	Adult 5+ yrs

ϕ

Age and sex specific ϕ and p

- *Robust estimates of lambda assuming stable age distribution*
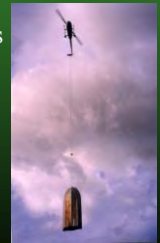
Owikeno Lake Case Study

- Declining sockeye salmon escapement impacting bear species (Hildebrand et al CJZ 1999)
  - Minimal ability of bears to switch to other food sources
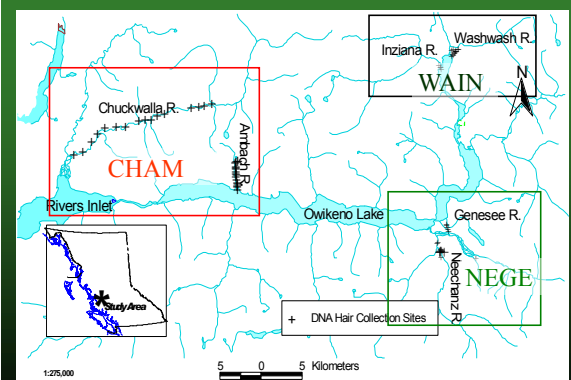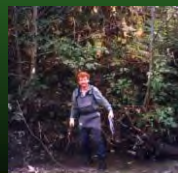

Monitoring challenges

- Remote and rugged study area
  - No road access
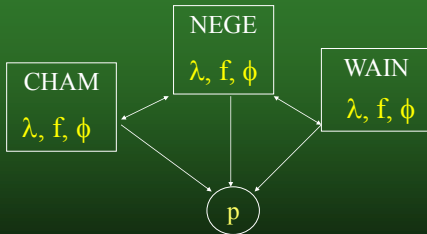- Suspected low sample sizes of bears

Methods

- Yearly sampling of 3 watershed areas during peak salmon escapement
- Barbed wire DNA sampling on bear trails adjacent to salmon streams
  - Pool yearly data
- Monitoring of salmon availability

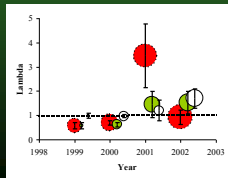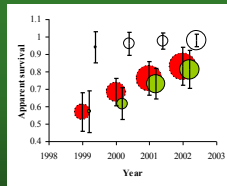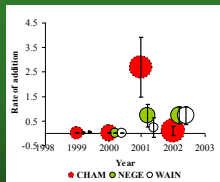## Pooling data to test hypothesis and increase precision

NEGE
λ, f, φ

CHAM
λ, f, φ

WAIN
λ, f, φ

p

• Salmon availability as a covariate to explain differences
• River-specific p versus effort curves

Model selection

⋙ River-specific capture probability versus effort curves

⋙ Salmon availability influences apparent survival and rates of addition

Demographic Analysis:

CJZ 2004

Owikeno conclusions

⋙ Multiple study areas explained temporal variation in trends

⋙ Covariates were quite useful
  • Salmon abundance
  • Effort

⋙ Demography influenced greatly by movements

Designing a trend study
A simulation approach

⋙ Population size has been estimated for most of the occupied habitat in Alberta

⋙ How can we estimate trend in the most cost efficient and informative way?

353
42
Edmonton
46
Calgary
90
51

Simulations to design trend project
Multiple data sources case study

⋙ HS-only grid projects are too expensive to repeat over many years

⋙ Can we combine data sources to reduce cost?
  • Hair snag as primary data set
  • Rub trees
  • Mgt bears and radio collared bears

Pradel model robust design

- Conduct multiple sessions within a year
  - Allows estimates of N each year
  - Allows estimates of trend (Pradel model)
  - Trend with 2 years of data
- Example data set (3 yr project-3 sessions/yr)
  - 010　　100　　000
  - 111　　000　　000
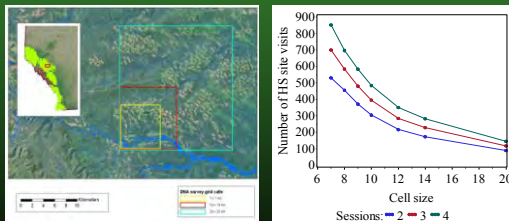
Multiple data sources
A less costly alternative?

Example: Collared (CO), Rub tree (RT), and HS sampling

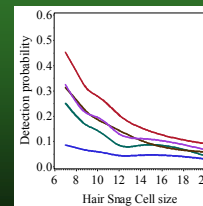| Year 1	Interval	Year 2
CO-RT-RT-HS-HS		CO-RT-RT-HS-HS
• Detection p • Population size	Apparent survival (θ) Additions (f) Population change (λ)	• Detection p • Population size
1011		0100

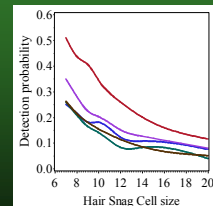## Reducing HS effort
Randomized resampling


Subsampling results

Females　　　　**Males**

─Castle　　─Grande Cache　─Yellowhead
─Clearwater　─Livingstone

Designs considered

Design	Cell size HS	Number of Sessions HS	RT	MG MF	Detection rates HS M	F	RT M	F	MG MF
HSRTMG (49)	49	2	2	1	0.38	0.28	0.1	0.05	0.1
HSRTMG (100)	100	2	2	1	0.19	0.16	0.1	0.05	0.1
HSRTMG (100),4x	100	4	2	1	0.19	0.16	0.1	0.05	0.1
HS only (49)	49	4			0.38	0.28			
HS only (100)	100	4			0.19	0.16			
RT only			4				0.1	0.05	
RT only (NDP)[a]			4				0.2	0.07	

Trends versus time
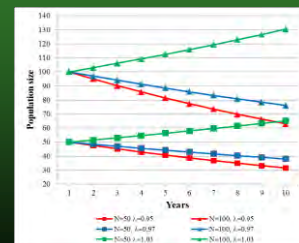
Comparison of designs
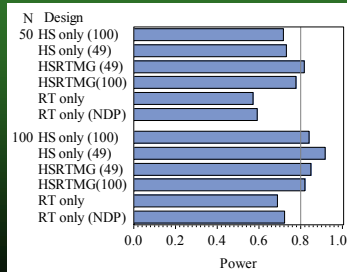
Power analysis

- Power depends on
  - Change in population, N and p
  - Hard to detect a small change quickly
- Adaptive calibration of effort as more data becomes available
  - Effort can be reduced after the first 3 years of annual sampling
  - Bi-annual, tri-annual sampling?
  - Simulations to further refine designs

Other mark-recapture methods

- Barker/Burnham models
  - Can utilize DNA, radio telemetry, and mortality data to estimate survival rates
  - Estimates reporting rate for all mortalities (hunting and other sources)
  - Refined estimate of survival
- Multi-state models
  - Estimate movements between 2 areas

DNA trend: Issues

- Age cannot be identified from DNA data
- Are "local demography"-based estimates adequate?
  - Multiple study areas
- Cannot determine exact causes of mortality
- Behavioral response to sampling?
  - Move sites between sessions to mitigate

DNA trends: Advantages

- Long-term skilled observers not needed
- Sample higher proportion of population
- Flexibility in modeling trends
  - Time varying parameters possible
  - Estimate process variance
  - Spatial partitioning of trend with multiple study areas
- Associations of trends with environmental/management factors

Conclusions

- Optimized study design is ESSENTIAL
- A team of biologists, geneticists, and statisticians should be used to design studies
- Mark-recapture analyses should complement previous study design and data collection
- Methods are still evolving……

Acknowledgements

Many people contributed to the research in this talk:

- Gord Stenhouse-Foothills Model Forest, Alberta
- Bruce McLellan, Ministry of Forests, BC
- Gary White- Colorado State University
- Mike Proctor-Birchdale Ecological
- Kate Kendall and Jeff Stetz-US Geological Survey
- Stefan Himmer-Arctos Consulting
- David Paetkau-Wildlife Genetics International
- Grant MacHutchon
- John Woods
- Many others…...